IRT Measurement Models for Conjoint Analysis

Justyna Brzezińska, Adam Sagan and Aneta Rybicka

Abstract Conjoint analysis and discrete choice models are widely accepted methods for preference measurement in marketing research. However, in all of these methods, the measurement of overall consumer preferences is based on binary, nominal or ordinal scales without implying any measurement model of these overall preferences. The aim of the paper is to propose Item Response Theory (IRT) latent variable models of overall preference measurement model for conjoint analysis. The model–based overall preference index (as a factor or ability scores) may be introduced into traditional conjoint analysis, instead of ordinal or choice-based preferences measured on weak scales without evidence of measure reliability. Two classes of models, Rasch-conjoint and nominal response-conjoint models, are developed and compared in the paper. The advantage of model-based preferences is to control for error of measurement and reliability (via standard

Justyna Brzezińska
University of Economics in Katowice, 1 Maja 50, 40-287 Katowice, Poland
✉ justyna.brzezinska@ue.katowice.pl

Adam Sagan
Cracow University of Economics, Rakowicka 27, 31-510 Cracow, Poland
✉ sagana@uek.krakow.pl

Aneta Rybicka
Wroclaw University of Economics, Nowowiejska 3, 58-500 Jelenia Gora, Poland
✉ aneta.rybicka@ue.wroc.pl

ARCHIVES OF DATA SCIENCE, SERIES A
(ONLINE FIRST)
KIT SCIENTIFIC PUBLISHING
Vol. 4, No. 1, 2018

DOI: 10.5445/KSP/1000085951/23
ISSN 2363-9881

error of measurement and test information function) of preference measurement model and the size of potential distortions related to preference scale unreliability and part-worth parameters bias. The comparative analysis based on the banking products described with 5 binary attributes was done on the sample of 542 respondents from 172 households in the southern part of Poland.

1 Introduction

Conjoint analysis consists of two broad classes of models: Traditional conjoint analysis (TCA), in which overall stated preferences are measured with weak rating (Likert-type), rank-ordered, pairwise comparisons and best-worse ordinal scale and choice-based conjoint (CBC), where overall preferences are revealed on the bases of several choice tasks.

All these approaches assume that stated overall preferences are measured without errors and Ordinary Least Squares, MONANOVA or rank-order logistic regression are used for the estimation of part-worth function parameters in the TCA, whereas various versions of multinomial logit models (conditional logit, alternative specific conditional logit, nested logit) are used in case of the CBC.

Lack of explicit measurement model for the identification of overall preferences makes the TCA and CBC problematic with parameter estimation that relies on a latent response variable concept as an approximation of continuous latent preferences. Second, the reliability of overall preferences cannot be estimated. There is no possibility of checking the size of parameters and standard errors attenuation in conjoint models due to the unreliability of the dependent variable.

To fill this gap we propose the IRT conjoint measurement model for measurement latent overall preferences. The Item Response Theory (IRT) models use binary, polytomous or nominal indicators of unobserved latent variables and create person ability scores as continuous measures of underlying measured personal traits. This enables to replace weak measurement scales and the CBC approach with an OLS regression-based model for conjoint analysis both in the TCA and the CBC framework. Introducing measurement error to overall preferences in an IRT (Rasch) model is related to a paradox known also as “Rasch paradox” (Michell, 2008a,b). The main idea is as follows: The Rasch model implies that eliminating all error factors from observational conditions dramatically decreases the precision of our observations in psychological testing.

This is paradoxical because it is normally thought that eliminating error factors improves the precision of observations.

2 Conjoint Analysis and IRT Models

2.1 Conjoint Measurement and Analysis

Conjoint measurement is simultaneous scaling (transformation) of the two variables, together with the response variable, to obtain an additive structure. Additive conjoint measurement using IRT models are models, where the item response probability is an additive function of two independent variables: (1) person ability and (2) item difficulty. Specific objectivity in the analysis is that person abilities do not depend on particular items and item difficulties do not depend on given persons. This property is given by a logistic response function. According to Coombs *A Theory of Data*, conjoint measurement is an Object × Person comparison (Coombs, 1964).

Conjoint analysis originated in mathematical psychology and was developed since the mid-sixties also by researchers in marketing and business. Conjoint analysis is a statistical method for finding out how consumers make trade-offs and choose among competing products or services. It is also used to predict (simulate) consumers' choices for future products or services. In our study IRT models are for TCA, however this techniques might be useful for CBC, Persona-Joint Method (PJM; Uchida et al, 2014) as well. Our approach is suitable for ordinal (paired-comparison, best-worst or ranking data) and CBC.

The main aim of conjoint analysis is to estimate part-worth utilities for attribute levels, as well as the ability to simulate all attribute (level) combinations to show the share of choice for different products. Part-worth utilities are estimated for each respondent separately and as average values for the whole sample. Estimated part-worth utilities allow to estimate the following values: Total utilities of the profile for all respondents, average total utilities in the sample, average attribute importance and average total utilities in the segments (clusters) of respondents. These elements are necessary for a comparative analysis of models.

A conjoint analysis model can be estimated at an individual level (number of models is equal to the number of respondents) and at an aggregated level. According to Coombs theory of data (Coombs, 1964), conjoint analysis is based on Object × Object comparison. However, in prevailing conjoint analysis studies, overall preferences of profiles are often measured on ordinal scales (i.e. Likert-type) without any assumptions concerning the measurement model (Steiner and Meißner, 2018).

2.2 IRT Models

Item Response Theory (IRT) is a statistical theory that distinguishes the latent trait (ability) of a participant from the difficulty of a set of items with well-correlated response patterns (Lawley, 1943; Rasch, 1960; Samejima, 1968; DeMars, 2010; Neubauer, 2003). IRT is a psychometric theory and a family of associated mathematical models that relate a latent trait of interest to the probability of responses to items on the assessment. It is a very general method, permitting one or more traits, various (testable) model assumptions and the analysis of binary or polytomous data. The mechanism of IRT can be presented most easily in terms of a dichotomous one-parameter (1PL) model, that is, a model for items with only two response alternatives. Such items require responses that are either correct or incorrect. In dichotomous IRT models, the item category that represents a positive response (and is subsequently coded 1) is described as indicating a correct response on an item; the alternative category, coded 0, indicates an incorrect response. Moreover, the item location parameter is commonly referred to as the item difficulty parameter. The IRT function is a function reflecting the conditional probability of selecting a positive response on an item.

The two-parameters (2PL) function requires the estimation of two parameters. One is a location parameter, which describes where along the trait continuum the function is centered. The second parameter is estimated to give an information on of how well an item discriminates among people along the trait continuum and shows how well an item can tell people apart with respect to the amount of a trait that they have.

The three-parameters (3PL) IRT model adds a guessing parameter that defines the non-zero lower asymptote of the item characteristic curve (ICC) that indicates above-zero item-response probability for extreme low person ability. And finally, the four-parameters (4PL) IRT model has an additional parameter that defines upper non-one asymptote of the item characteristic curve (ICC) that indicates below-one item-response probability for extreme high person ability.

Table 1: Classification and restrictions of IRT measurement models.

IRT Model	**Response Scale**	**Restrictions**
Nominal Response Model	Multiple choices	No restrictions
Partial Credit Model	Likert scales	Order restrictions
Samejima IRT Models	Binary responses	Boundary restrictions
Rasch model	Binary responses	Equality restrictions

A Nominal Response Model (NRM; Bock, 1972; Baker and Kim, 2004) is a model for nominally scored responses (items) that may vary in their difficulty and discrimination parameters. The item parameters are calculated using a multinomial logistic model. The difficulty parameter measures the propensity to choose a given item category instead of the base outcome (the first item usually). The probability of responding in any given category is modelled directly by multivariate generalization of the logistic latent trait model. The NRM is an extension of the multinomial, fixed-effects, logit-linear models to the mixed-effects setting of IRT. The implicit purpose of the NRM is to find the unobserved ordering within unordered categorical data. We present classification and restrictions of IRT measurement models in Table 1.

3 Research Design

The research on banking products and bank account preferences was conducted in Poland. The product analyzed was bank account choices of bank customers. The questionnaire consisted of the following yes/no questions for the following 5 attributes:

1. Bank account access via mobile devices (X_1),
2. bank account commission (X_2),
3. credit card payment return (X_3),
4. fee for withdrawal in foreign ATM machines (X_4), and
5. credit card free of charge (X_5).

A full factorial design contained 32 profiles (depending on the number of levels and variables in the study: $2 \times 2 \times 2 \times 2 \times 2 = 32$). However, a fractional factorial design was developed and prepared using R programme and the final design consists of 8 profiles. As a result, the respondents (231 in total) were asked to make a choice between 28 pairs of profiles. The profiles for the Rasch and NRM models of banking products are given in Table 2.

Table 2: Profiles in conjoint analysis.

Profile	X_1	X_2	X_3	X_4	X_5
A	no	no	yes	yes	yes
B	yes	yes	no	yes	yes
C	no	yes	yes	no	yes
D	yes	no	no	no	yes
E	yes	no	yes	yes	no
F	no	yes	no	yes	no
G	yes	yes	yes	no	no
H	no	no	no	no	no

3.1 Measurement Rasch Models

The Rasch model is the simplest one parametric IRT model assuming equality of discrimination parameters (discrimination parameter = 1):

$$P(Y_j = 1 \mid \theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)} . \tag{1}$$

The Rasch model of preferences was estimated using a structural equation model on paired comparison data (Maydeu-Olivares and Böckenholt, 2005). The structural model of pairwise comparison of profiles is based on latent utilities model where the choice between i and k profiles is given as:

$$y = \begin{cases} 1 \text{ if } t_i \geq t_k \\ 0 \text{ if } t_i < t_k \end{cases} \tag{2}$$

where t_i and t_k represent the latent utilities of i and k profiles. The response process can be described as differences between latent utilities $y_l^* = t_i - t_k$. And, therefore:

$$y = \begin{cases} 1 \text{ if } y_l \geq 0 \\ 0 \text{ if } y_l < 0 \end{cases} \quad (3)$$

The response process in matrix form can be rephrased as $y^* = At$, where y^* is the vector of latent difference responses, A is design matrix and t is vector of latent utilities.The design matrix is based on pairwise comparisons of all n profiles and has the form $\frac{n(n-1)}{2} \cdot n$. The CFA model for pairwise data is set of linear relations of $\frac{n(n-1)}{2}$ indicators y^* on p latent variables η defined as $y^* = \nu + \Lambda\eta + \epsilon$, where ν represents the $\frac{n(n-1)}{2}$ dimensional vector of the intercepts for the measurement equation, λ is an $\frac{n(n-1)}{2} \times p$ matrix of factor loadings (1 and -1), and ϵ is $\frac{n(n-1)}{2}$ dimensional vector of residuals.

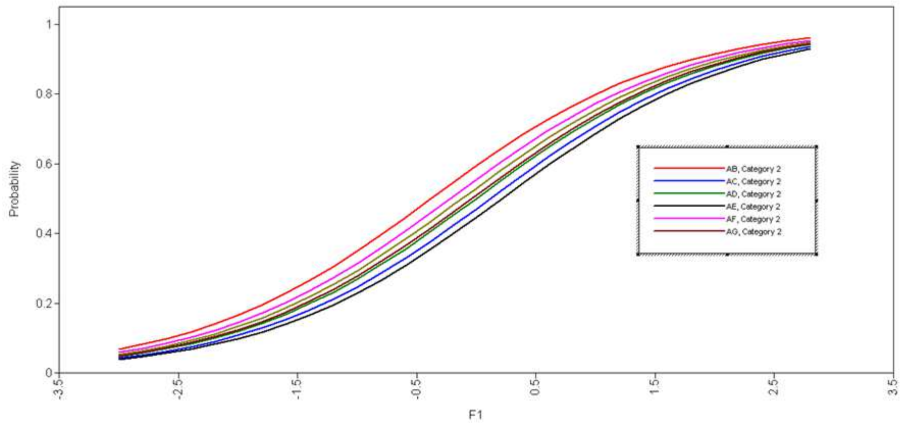

Figure 1: Rasch model (profile A).

Specific items represent preferences of profiles (A over B, A over C etc.). Factor loadings are fixed to 1 (if A dominates B) or -1 (if B dominates A) for particular items. The means of latent preference for the profiles are fixed to 0 and thresholds for binary items are estimated. The model assumes unidimensionality of preferences. The number of models is equal to the number of profiles. Only positive domination structures are shown on the item characteristic curves. Preference analysis is based on the comparison between appropriate pairs of items. For example, the comparison of the AC threshold (0.118) in profile A model with the CA threshold (−0.112) in profile C model, gives the information about the preference C over A, whereas the comparison of AD (0.001) with DA (−0.112) results in the preference of A over D etc.

Table 3: The parameters of the Rasch models.

Profile	.A	.B	.C	.D	.E	.F	.G	.H	χ^2	df	p
A	- - -	−0.380	0.118	0.001	0.217	−0.219	−0.041	−0.120	300.22	120	0.000
B	0.401	- - -	−0.362	0.259	−0.021	−0.321	−0.548	−0.465	211.22	120	0.000
C	−0.112	0.370	- - -	0.288	0.088	−0.293	0.188	0.028	182.12	120	0.000
D	−0.112	0.370	0.288	- - -	0.088	−0.293	0.188	0.028	182.12	120	0.000
E	−0.202	0.035	−0.083	0.174	- - -	−0.302	−0.322	−0.282	189.34	120	0.000
F	0.241	0.343	0.302	0.322	0.322	- - -	0.261	0.017	168.14	120	0.020
G	0.056	0.559	−0.182	0.155	0.335	−0.242	- - -	−0.362	234.04	120	0.000
H	0.120	0.465	−0.039	0.465	0.281	−0.019	0.362	- - -	220.4	120	0.000

The item characteristic curves for the profile A are depicted in Figure 1. The ICC shows the conditional probabilities of endorsing profile A. The lower (reading row-wise) the difficulty (location) parameter is, the higher the location of the ICC curve and the probability of responding to specific preferences of the A profile with respect to the others. The parameters of the Rasch models are presented in Table 3.

To summarize, the Rasch-based measurement model enables to use estimated factor scores as a metric latent variable of banking product preferences for a conjoint model.

3.2 Nominal Response Measurement Model

Now we present a nominal response measurement model (Table 4). The model was estimated on the basis of the overall choices between eight profiles. the nominal response model or nominal category model is based on the multivariate logit model where for item i, the probability of person j choosing category k on item i, given person j latent trait θ_j is (Bock, 1972; Baker and Kim, 2004):

$$P(Y_{ij} = k|\theta_j) = \frac{\exp(a_{ik}(\theta_j - b_{ik}))}{\sum_{h=1}^{K} \exp(a_{ik}(\theta_j - b_{ik}))} \tag{4}$$

where a_{ik} represents the discrimination of category k for item i, b_{ik} is the difficulty (location) of category k for item i, and θ_j is the latent trait of person j. The first outcome category is the base outcome with which other parameters will be compared; this implies the constraint $a_{i1} = 0$ and $b_{i1} = 0$ for each item i. With this constraint, a_{ik} and b_{ik} are the discrimination and difficulty to choose category k relative to the first category. Profile A was selected as a reference profile for the preference analysis.

Table 4: Model parameters.

Profile	Discrimination Parameters	Difficulty Parameters
B	0.90	−0.29
C	0.85	−0.77
D	0.86	−0.69
E	0.91	−0.08
F	0.76	−1.43
G	0.94	0.25
H	0.85	−0.77

The difficulty parameters of B-H profiles are related to the difficulty of base profile A ($b_1 = 0$) and also the discrimination parameters are related to the discrimination parameter of base profile A ($a_1 = 0$). An item with a large discrimination value has a high correlation (with respect to base profile A) between the latent trait and the probability of success on that item. In other words, an item with a large discrimination parameter can distinguish better between low and high levels of the latent trait. We present category characteristic curves in Figure 2 and boundary characteristic curves (BCC) for the nominal response model in Figure 3.

In the nominal response model the category characteristic curves (CCC) show the probability of endorsing a particular category (here profile) given the latent trait (theta). We see that respondents with the latent trait level below approximately −1.7 tend to endorse profile 1 (blue, A), and respondents with a latent trait level above that point tend to choose profile 6 (red, F). BCCs represent cumulative probabilities and represent the theta point (value on the latent preference dimension for the person) where the probability of responding in profile *k* or higher is 0.5. We present the test characteristic curve in Figure 4 and the total information curve for the nominal response model in Figure 5. The test characteristic curve is the sum of ICCs for the entire instrument (all profiles) and thus plots the expected score on the preference scale along the latent trait continuum.

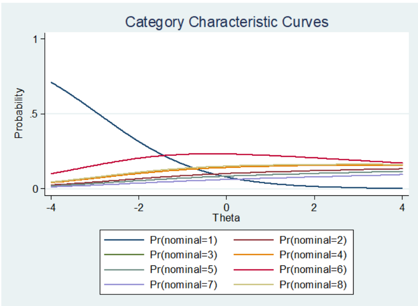

Figure 2: Category Characteristic Curves (CCC).

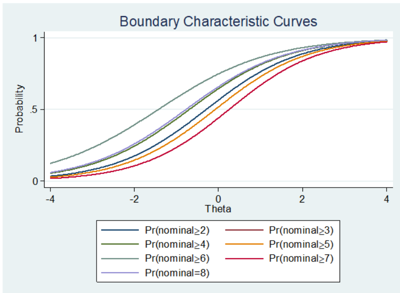

Figure 3: Boundary Characteristic Curves (BCC).

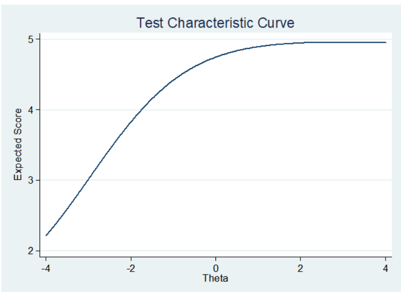

Figure 4: Test Characteristic Curve (TCC).

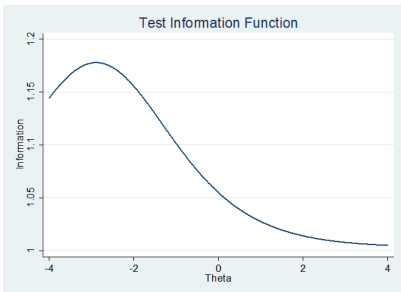

Figure 5: Total Information Curve (TIC).

In IRT, the term information is used to describe the reliability or precision of measurement along the latent preference (theta). More reliable instruments measure the latent trait around the estimated difficulty parameter with greater precision. The TIC, displayed in Figure 5, shows that the preference scale is more reliable at the left extreme end of the scale (lower preferences) and less reliable for average respondents and those with higher preferences. On the basis of the IRT models the factor scores (theta) for all 8 profiles are obtained and can be used as dependent metric variables for regression-based conjoint analysis. Because of the measurement error (unreliability) of IRT, preference scales are incorporated into the conjoint model, the part-worth parameters are therefore corrected for attenuation.

To sum up, a Rasch model was estimated using pairwise comparisons between items (A–B), (A–C), (A–D) etc. so it reveals preferences between pairs of profiles. The nominal response model is related to the choice between the items and is estimated using a multinomial conditional logit model that enables overall comparisons with the reference profile (profile A).

4 Conclusions

The presented Rasch and nominal response IRT models for conjoint analysis allow the introduction of a measurement model in the estimation of overall latent preferences of product bundles (profiles) for conjoint analysis and to overcome the problems with weak ordinal scales for preference measurement in traditional conjoint models. IRT-based analysis introduces the measurement model for overall preference measurement, whereas in classic Conjoint Analysis no measurement model is used for measuring the overall preferences.

The IRT-based approach helps to assess the amount of information and error of the measurement in the preference scale and, therefore, the reliability of scale and possible corrections for attenuation of parameter estimates. Correction for attenuation of parameter estimates deals with the problem of the influence of reliability on regression coefficients. The lower the reliability, the stronger the attenuation of “true” regression (parth-worth) coefficients in conjoint analysis.

The use of the IRT model allows replacing a weak profile measurement scale with a scale with known psychometric properties. Having an overall preferences’ scale as a single dependent variable with known distributional properties and reliability, the part-worth functions can be estimated using a classic metric conjoint model and they can be used for preference simulation.

References

Baker FB, Kim SH (2004) Item Response Theory: Parameter Estimation Techniques, 2nd edn. CRC Press, Taylor & Francis, Boca Raton (USA). ISBN: 978-0-824758-25-7.

Bock RD (1972) Estimating Item Parameters and Latent Ability When Responses Are Scored in Two or More Nominal Categories. Psychometrika 37:29–51, von Davier M, Ip E, Marsman M (eds). DOI: 10.1007/BF02291411.

Coombs CH (1964) A Theory of Data, 1st edn. John Wiley & Sons, New York (USA).

DeMars C (2010) Item Response Theory. Oxford University Press. ISBN: 978-0-195377-03-3.

Lawley DN (1943) On Problems Connected with Item Selection and Test Construction. Proceedings of the Royal Society of Edinburgh Section A: Mathematics 61(3):273–287. DOI: 10.1017/S0080454100006282.

Maydeu-Olivares A, Böckenholt U (2005) Structural Equation Modeling of Paired-Comparison and Ranking Data. Psychological Methods 10(3):285–304. DOI: 10.1037/1082-989X.10.3.285.

Michell J (2008a) Conjoint Measurement and the Rasch Paradox: A Response to Kyngdon. Theory & Psychology 18:119–124. DOI: 10.1177/0959354307086926.

Michell J (2008b) Is Psychometrics Pathological Science? Measurement: Interdisciplinary Research and Perspectives 6(1-2):7–24. DOI: 10.1080/15366360802035489.

Neubauer G (2003) An IRT-Approach for Conjoint Analysis. Development in Applied Statistics, Metodoloski zvezki 19:35–47, Ferligoj A, Mrvar A (eds).

Rasch G (1960) Probabilistic Models for Some Intelligence and Attainment Tests. Proceedings of the Royal Society of Edinburch Copenhagen.

Samejima F (1968) Calibration of Latent Ability Using Response Pattern of Graded Score. ETS Research Bulletin Series 1968(1). DOI: 10.1002/j.2333-8504.1968.tb00153.x

Steiner M, Meißner M (2018) A User's Guide to the Galaxy of Conjoint Analysis and Compositional Preference Measurement. Marketing ZFP 40(2):3–25. DOI: 10.15358/0344-1369-2018-2-3.

Uchida H, Kunigami M, Terano T, Yoshikawa A (2014) Persona-Conjoint Method to Measure Learners' Changes of Implicit Understanding in Business Situation. In: Proceedings of the 6th International Conference on Education and New Learning Technologies (EDULEARN14), International Academy of Technology, Education and Development (IATED), Valencia (Spain), pp. 6942–6949. ISBN: 978-8-461705-57-3, URL: `https://library.iated.org/view/UCHIDA2014PER`.